DOCUMENT RESUME

ED 340 741 TM 017 803

AUTHOR Bogotch, Ira E.; Taylor, Dianne L.
TITLE Discretionary Assessment Practices.
PUB DATE Nov 91
NOTE 38p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (20th, Lexington, KY, November 12-15, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Administrator Attitudes; Administrator Role; Classroom Observation Techniques; Decision Making; *Educational Assessment; Educational Practices; Elementary Secondary Education; *Evaluation Methods; Interviews; *Principals; Standardized Tests; *Teacher Evaluation; Videotape Recordings

ABSTRACT

Descriptive statements of evaluation practices by principals were studied concerning normative expectations and J. Dewey's general logic of experience. Focus was on describing what principals think and do in terms of evaluation as a step toward bringing evaluation specialists and educational practitioners into a closer working relationship. Four elementary school and two secondary school principals from a large urban school district were videotaped while talking about their experiences. All of the subjects referred to standardized tests as the primary measure used to evaluate academic achievement. Although they described standardized tests as a necessary evil, principals often pressured teachers to teach subject matter in the way it is to be tested. Principals described several qualitative evaluation techniques; focused on aspects of education beyond academic excellence, such as children's self-esteem, cultural diversity, and extracurricular activities; and recognized that evaluation, while subjective, involves trying to understand the total school context on a daily basis. Although gaps exist concerning the logic and coherence among processes and indicators by Dewey's standards, little evidence of final judgments and opinion-centered evaluation was found. Principals do not believe that formal evaluations tell them much about teachers, yet they must rely on them. Evaluation now occurs haphazardly and informally. Recommendations are made for the improvement of principals' evaluation decisions. One table and two figures present data from the study. A 23-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340741

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

IRA E. BOGOTCH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Discretionary Assessment Practices

Ira E. Bogotch
University of New Orleans

Dianne L. Taylor
Tulane University

Paper presented at MSERA, Lexington, KY, November, 1991

TM017803

BEST COPY AVAILABLE

Discretionary Assessment Practices

Popham (1988) establishes the conventional boundary for educational evaluation claiming that "all the fuss in education about evaluation is not focused on ...informal evaluative acts," but rather on the "systematic" or "a formal appraisal of the quality of educational phenomena" (p. 7). Such a normative view hardly touches the on-going, daily evaluative activities of school practitioners, except, perhaps, as threats to the latter's competency. The gap between evaluation theory and practice is twofold. Not only are evaluative theory and research ignored by decision-makers and policy makers, but both have had only minimal impact on what practitioners -- administrators or teachers -- regularly do. The center of controversy - at least with respect to accountability - may be at the level described by Popham, but it is on the level of action (Vickers, 1967) and critical judgment (Dewey, 1939/70) where theorists and practitioners need to join together to improve evaluative behaviors and judgments.

The underlying assumption here is that evaluations must be conducted within the total experiences and activities of practitioners. The world of experience is so complex that without a general logic of experience, improving practice is impossible. School practitioners are often accused of not learning from their practice or from each other. They are also accused of being atheoretical, that is, not able to analyze what is happening beyond a particular time- and space-bound context. To a great extent, the daily administrative tasks are difficult

and confusing enough (**Mintzberg**, 1971); but, that limited perspective does not build professional practice, improve schools, nor speak to the issues of professional and public trust.

Historical Milestones in Educational Evaluation

The history of educational evaluation is well-known to specialists in the field. It unfolds in discontinuous eras marked by shifts in direction which often seem to reject the previous focus. The early signposts included the development of tests measuring student achievement [e.g., Stanford] and intelligence [e.g., Binet] followed by comparative analyses describing the strengths and weaknesses of programs based on clearly stated goals [e.g., The Eight Year Study]. The next era incorporated judgment into educational evaluation and is characterized by a number of persuasive, although disparate, articles by Cronbach (1963), Stake (1967), and Scriven (1967).

Another phase is marked by a variety of rational models combining description and judgment with decisions. The resilience of these rational models can be explained primarily by the requirements of sponsored research (Berke & Law, 1981). Work published by the Joint Committee on Standards for Educational Evaluation (1981; 1988) represents the latest establishment position; while evaluation theories based on social constructivist views of reality requiring the shared participation of stakeholders (Guba & Lincoln, 1989) is the most recent paradigm shift.

These turns in direction reflect a wide range of views on evaluation; yet, with very few exceptions, the history of educational evaluation is based on the narrow perspectives of trained specialists as opposed to that of school practitioners. In other words, it emphasizes **what evaluators should do** rather than **what practitioners actually do** to evaluate programs and personnel. In this sense, it is reflects Bailyn's (1960) critique that histories of education are written by professional educators and, therefore, are limited histories which miss the fuller context of educational experiences.

Rather, the key to school improvement is through the quality of our professional judgments. Quality is necessarily a philosophical concept [e.g.'s Aristotle's model of logical categories and Kant's categories of knowledge) related to knowledge; neither rational technology nor a logic of confidence (Meyer & Rowan, 1977) can fully account for professional behaviors based on critical judgments. The former ignores the qualitative dimension of values, intuition, and judgment (viz., Kant); while, the latter tends to hide the subtle, contextual experiences of school building administrators. Instead, Dewey's general logic of experience represents an alternative framework for describing and analyzing school-level assessment behaviors.

The purpose of this study is to discuss the descriptive statements of evaluation practice in terms of both normative expectations and Dewey's general logic of experience. A descriptive analysis of what practitioners think and do in terms

of evaluation is a necessary, albeit preliminary, step towards bringing the evaluation specialist and the practitioner into a closer working relationship. When evaluation designs are linked more to external accountability requirements than to local educational realities and school improvement, the task becomes a ritual and cannot be very rewarding professionally. Likewise, non-reflective practice consigns practitioners to repeat past errors continually and suffer both deserved and undeserved indignities. Hence, it should be in the best long-term interests of both parties to work together. A more immediate objective, however, is to encourage professional scrutiny and public testing of what is actually happening within schools.

Dewey Revisited

Dewey's concept of reflective thinking provides a logical framework for viewing the historical events highlighted in the first paragraph. In distinguishing between science and philosophy, Dewey (1939/1970) states that science measures the facts of the world, whereas philosophy critically judges the totality of the world in-action (p. 182). His epistemology goes beyond measuring physical facts and making objective, quantitative comparisons. Such activities certainly are useful for critical thinking, but they are not judgements. They do not encompass "the full range of existence and experience" (Elliot, 1972) as Dewey stated. Critical judgement is based on the notion that experiences are on-going, they are neither complete nor final (p. 181). Therefore, evaluation "occurs whenever a moment

is devoted to looking to see what sort of value is present..., whenever we raise even a shadow of a question about its worth, or modify our sense of it by even a passing estimate of its probable future" (p. 184).

The data of critical judgments are perceptions, including those of politics and history. These dynamics add complexity and confusion to an already complex school work environment. What is needed, therefore, are individuals who are capable of discriminating and unifying a complex reality. The method which Dewey proposes is based on a general logic of experience. It begins with perceptions of conduct, beliefs about situations, and "appreciative perception which are characterized by immediate qualities of good and bad" (p. 188-89). The goal is to support whatever is good within experience by making it more coherent, more secure, and more significant.

The method for this general logic of experience is qualitative. It requires intelligent individuals with rich past experiences and disciplined insight capable of creating a clearer understanding of an action in relationship to its total context. The method involves hypothesizing, asking questions, modifying beliefs, searching for new facts, and choosing alternatives. The general logic of experience is a process model as depicted in **Figure 1.**

Figure 1. General Logic of Experience Process

Intelligence and Rich Backgrounds [Politics & History]

Perceptions (Measures & Data) [Qualities, Conduct, Beliefs]

Asking Questions
Hypothesizing
Modifying Beliefs
Selecting New Facts
Choosing Alternative Facts

Critical Judgements [On-going, Incomplete, Immediate, but not Final]

Discriminate and Unify Experiences

Totality [More Coherent, More Secure, Support of Good]

Dewey argues that this method of inquiry is superior to final judgments, opinions, and impressionistic criticism. The first seeks to settle a matter arbitrarily, the second is merely arbitrary, while the third leads to the chaos of subjectivity. For Dewey, as well as Whitehead (1951), the subject matter of education is based on the active qualities of living and thinking. It is on-going and contextual, not static, inert, or made up of isolated facts. However reassuring final judgments are to "unregenerate" minds (Dewey, 1939/1970, p. 185), critical judgments need to be tentative, hypothetical, and supportive of good work. Neither formalistic nor objective evaluations lend themselves to this kind of epistemology.

The questions raised by Dewey's work in contrast to that of the evaluation specialists are (a) whether individuals responsible for evaluation possess the necessary level of discriminating intelligence to make good judgments, (b) whether there is an epistemology embedded in our practicing school evaluative processes, and (c) whether evaluative decision-making and choice follow a logical and psychological method (Mintzberg, Raisinghani, & Theoret, 1976; Simon, 1946). The first question relates to selection and training of practicing school evaluators and is not addressed in this study. The question of epistemology of evaluative practice, however, is considered in terms of describing and categorizing what school principals think about and profess to do as part of their evaluative activities. In terms of the third question about decision-making, which involves

judgment, analysis, and bargaining (Mintzberg, et al., 1976), evaluation must extend beyond a cerebral knowledge about evaluation theory and practices and become an experientially grounded knowledge concerning the conduct of evaluation in educational settings (Ryle, 1969).

Purpose of the study

This empirical study seeks to describe what practitioners think and do when evaluating. It is a search for language, meaning, and action. In two earlier studies, it was reported that both principals and teachers agree that (a) structural evaluations have less relevancy and frequency than discretionary activities, (b) these discretionary activities are neither less significant nor incidental to improving schools, and (c) there is a statistically significant difference between discretionary activities identified by the qualities of fairness and worth as compared to the structural-frequency of school evaluations processes (Bogotch, 1990; Williams, 1991). The findings demonstrated that discretion was exercised along a continuum, at times linked closely to structural mechanisms, while at other times, discretion were completely idiosyncratic. The present study explores these relationships further using the experiential framework of Dewey's reflective thinking. The question is whether there exists a "structure" or logic underlying on-going evaluative judgments. In pursuit of this objective, the evaluation specialists' knowledge is useful, not as generalizable theories or behavioral models (Guba & Lincoln, 1989), but rather,

as descriptive categories of evaluative analyses and actions.

Method and Procedures

Subjects:

Two groups of principals (N=6) from a large urban school district were invited to videotaped interview sessions. The first group included four elementary school principals whose school based administrative experience ranged from 3 years to 12 years. The combined faculty total at their schools was 153 teachers. An additional characteristic of importance is that these four principals were enrolled in educational administration doctoral studies at the time of the interviews.

The second group consisted of two secondary school principals: one from a middle school, grades 7 and 8 and the other from a secondary school, grades 7 through 12. The combined faculty at these secondary schools was 113 teachers. Their experience as school principals totalled 21 years.

Procedures

More precise language is needed in order to describe what principals think and do as part of evaluation of programs and personnel. The conception of evaluation as on-going growth processes involves many incomplete thoughts, intangible ideas, and diverse behaviors such that attaching words to these processes is not easy for practitioners. Each principal brings his or her specific ideas and the language to express them. In a focus group setting, however, listening to how others describe what they do may trigger new language. Therefore, principals

were brought together by instructional levels to talk as a group about evaluation practices.

The videotaped sessions were held at a nearby state university campus on two separate afternoons. Each session followed the same interview protocol of questions and lasted 75 minutes. The researchers viewed the tapes together, coded the language used by the principals, and categorized responses in terms of the critical judgment framework described by Dewey in **Figure 1.**

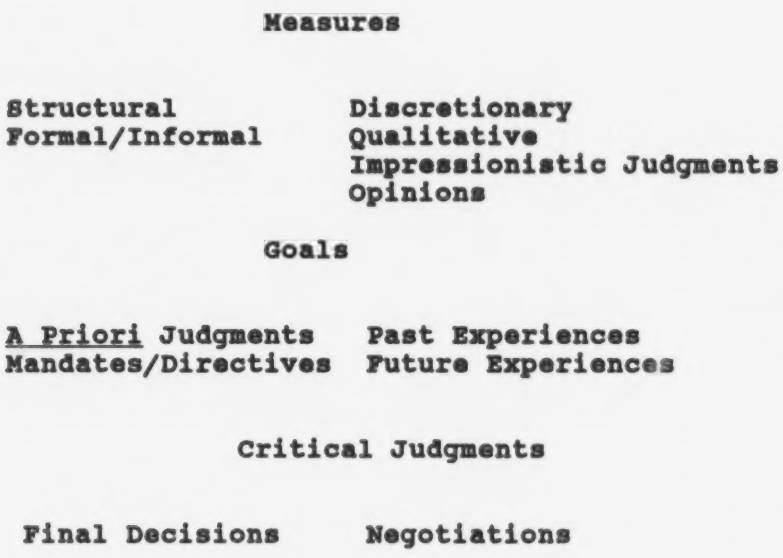
The statements of the principals are taken at their face value. That is, we have recorded what the principals said they did as well as what they said their actions meant. We feel compelled state this up front because the principals admit performing evaluation activities based on real world assumptions. Within this political context, there are institutional constraints to telling the truth (McDermott & Tylbor, 1987). Future studies need to be directed towards this social phenomenon. Finally, no one study using a single method with one population sample can claim validity. If this line of research is to contribute to understanding evaluation, others must conduct further studies.

Results

Beyond the list of formal evaluative mechanisms and structural constraints - cyclically mandated formal evaluations of teachers, programs, and textbooks; due process procedures; union contracts; etc - school principals do not find it easy to

articulate what they actually do when they evaluate people and programs. There are so many intangibles which are part of the total picture that simplistic procedures, indicators, or guidelines fail to capture the continuity and interactions within schools on a daily basis. Despite this complexity, principals say they know instructional and curricular quality when they see it. The data reported here are indicative of the many discretionary and qualitative ways by which principals measure, judge, decide, and negotiate evaluation within their schools. **Figure 2** represents one conceptual schema based on the data from this study.

Figure 2. Interactions Among Objective and Internal Conditions



Interactions between Structural and Discretionary Measures

All of the principals referred to standardized student test scores as the "primary" measure used to evaluate academic achievement. It is the "published reality" by which policymakers and the public judge not only school improvement, but also whether teachers and administrators are doing a good job. In other words, standardized test scores qua measures are elevated by non-practicing educators to the status of educational goals. As a result, curricular decisions made by state and central office administrators are tied to standardized tests. To principals, this measure in "unfair," "unrealistic," and "overly simplistic." It is viewed by principals as an "obstacle" to evaluating both programs and instructional personnel. To one secondary principal, equating test results with good or bad teaching is tantamount to deducing that "one and one equals four." Rather than viewing test results as evidence of student academic growth or of whether a teacher "is in the ballpark," the tests drive instructional techniques and curricula. Using test results for these purposes "sells all of us short," according to this principal.

Although principals view standardized tests as a "necessary evil," teachers are often pressured by principals to teach subject-matter in the exact way it is tested. For example, one elementary school principal investigated low test scores in spelling and found there was a poor match between the words tested and those taught through the approved curriculum.

Teachers at this school also noted that the testing procedures used in the classroom differed from those used during testing. The principal pressed for the abandonment of traditional spelling test procedures, whereby the teacher says the word aloud and students write the word on paper, in favor of the procedure used on the standardized test, in which students selected the correctly spelled word from among four choices. In this instance, the test dictated pedagogical decisions as to appropriate teaching practices. Such instances notwithstanding, principals did not feel as if they were victims of standardized testing since they take test scores into account as part of the total curricular and instructional picture.

A larger view of evaluation emerged from a seemingly unsystematic list of behaviors and perceptions offered by the principals. Principals tend to refer to these as informal behaviors, and not as evaluation **per se**. Yet, such behaviors are precisely the on-going evaluative activities which contribute to judgments, decisions, and negotiations. In previous studies, these behaviors were referred to as discretionary assessment behaviors (Bogotch, 1990) and as quality control (Williams, 1991). Here, however, it seems more appropriate to label them as qualitative measures since their function is as data for further evaluations within a broader contextual setting. The comparative data from the two studies are summarized in Table 1.

Table 1:

Discretionary, Qualitative Assessment Measures

Bogotch (1990)	Bogotch & Taylor (1991)
Sit down informally	Regularly enter classes
"Tell me what's happening."	Ask "what's happening?"
Talk about positive things	Stand in the hall
We prioritize one or two aspects.	Listen to classes
The instrument leaves us some discretion.	Appropriate noise levels
Negatives are better to say face to face.	The kinds of questions asked in class
Teachers who need information will seek it.	Feedback from parents, students, teachers
Periodically attend grade meetings.	"Are students treated with respect?"
Review mid-period reports.	Consultations with staff members [librarians, department heads, social workers, special ed teachers,
Look at samples of students' work	Review lesson plans
Listen to children read	Review field trips
See how often children pick up a book	Review number and type of student referrals
Everyone seems to like it.	Requests for guest speakers
What's the demand for next year?	Requests for money
Feedback from students and teachers	Student and parent complaints
Watch students	Quarterly exams
Check computer logs	Sponsoring of clubs
[Evaluation] criteria is discussed	Waiting lists for programs
Grade level consensus	

Among the qualitative measures described by the principals in the present study were: regularly entering classrooms; asking teachers, students, and parents "how everything is going;" standing in the hall and listening to classroom interactions, particularly to the cognitive level of the questions being asked and to the appropriateness of noise; getting feedback from parents and students about programs and personnel; determining whether children were treated with respect; consulting with other staff members including the media specialist, librarian, guidance counselors, department heads, social worker, special education teachers, and resource teachers on the use of books, media, and information; reviewing lesson plans, field trips, the number and type of discipline referrals made by teachers, quarterly exams, requests for guest speakers, requests for money and materials; listening to parent and student complaints; and considering which teachers sponsor clubs. Demand is also viewed as a qualitative measure; therefore, a program or school that has a long waiting list or a high enrollment is evaluated positively since "people want to be part of a good program."

Some of the principals labelled these qualities as "extras;" citing both individuals and schools where "extras" were not evident or emphasized, and yet were considered excellent. Others felt these qualities were associated with good instruction and worthwhile programs. For all principals, however, the criteria of effort and preparation were viewed as correlates, not causal factors, of good schools.

Negative counterparts to many of the above qualitative measures were also expressed. Among these negative criteria were inappropriate noise, lack of control, uninteresting lessons, inappropriate and inaccurate classroom presentations, and embarrassing children.

The formal-qualitative dichotomy also emerged during the discussion of formal instructional evaluation instruments. On the positive side, these formal instruments and procedures were said to help principals refocus on specific teaching acts, some of which they themselves might have otherwise ignored. But as far as providing truthful "objectivity" in terms of a total picture, principals felt that formal instruments fall far short. "We have failed to develop an instrument which measures all that a teacher is supposed to do." The gap between the mandated number of teaching behaviors recorded during an observation and the on-going context for teacher evaluation is apparent in the comments made by the principals. To these practitioners, any one classroom lesson represents a snapshot along a continuum. Principals do not expect to see all of the mandated behaviors simply because it may not be relevant to that aspect of the learning continuum for that class at that particular time.

Even if the principal were not consciously aware of the long-term objectives for a class, the realities of teaching in a particular cultural environment necessarily shapes teaching behaviors. Thus, the requirements established by a uniform checklist ignore the realities that principals and schools have

instructional priorities which are more valued than generic behaviors (Bogotch, 1990). "The sum of the parts do not equal the whole when it comes to evaluating teachers." "No two schools are alike." Neither are any two teachers.

Interactions between Mandates and A Priori Judgments and Experiences

The consensus among principals was that the public has expectations that are not consistent with the realities of schools. One principal described these external expectations as relative to what is known and done in noneducational contexts. According to this principal, many tasks performed by parents and business people involve educational functions; therefore, non-professional educators feel qualified to discuss professional judgments. To circumvent this problem, the principal invited parents to the school so that they could become familiar with its unique environment. In his estimation, most outsiders are impressed and typically comment, "I don't know how you do what you do."

To the public, what principals do in terms of evaluation is a mystery. In part, one principal explained, this is because very few individuals remember having contact with the principal during their years as a student or seeing principals evaluate teachers and programs. Yet, principals define their roles quite clearly in non-technical terms: "I basically follow a model. I decide on a focus and vision, such as to improve academic achievement." "I ask myself, 'Did I create an environment where

students know what to do and teachers know where they are headed?'" "Does classroom instruction exceed minimum requirements?"

While academic excellence often was stated as a primary focus, principals also focus on other aspects of education such as the whole child, children's self-esteem, cultural diversity, social clubs, performing groups, and sports activities. "We try to provide diverse experiences for children. There are kids who live in the city who have never been to the zoo."

These statements, while important, do not present a complete picture in terms of goals and objectives. Principals mentioned both state and system mandated guidelines for curricula. Yet, in terms of curricula evaluation, the most common principal response was that they were not curricular experts, and so relied on the expertise of the teachers. If there were no teacher experts on the faculty, the principals would either seek outside models as resources or make it a priority to improve expertise in a particular grade level or department. Statements of trust, faith, and reliance upon teachers were unanimous; "I trust the expectations of the teaching staff," was typical. With regard to curricular evaluation, however, others also noted that "There are a lot of intangibles, especially when we try to articulate them." For example, one principal commented, "I cannot tell you why, but if a teacher says that this is or is not working, I trust that."

A part of this faith in teachers as curricular experts was based on the belief that subject-matter issues - mandated by

states and districts - were not of primary concern within their schools. The principals voiced trust not only in university teacher education programs, but also in the state and district guidelines, texts, and the many published resources which are available to any teacher who needs further information on a curricular topic. Although principals mentioned that they check for inaccuracies in lessons, this was not perceived as significant a problem as was instructional improvement.

Nevertheless, the trust placed in teachers was qualified. For example, teachers, especially in elementary schools, need to be monitored so that they do not teach just what they like, according to the elementary principals. Likewise, a high school history teacher who devotes a significant portion of the academic year to the Civil War was also perceived as not doing an adequate job. In this sense, the principals looked for "a balance" in the curricula.

Focusing on their curricular role, elementary school principals expressed the need to be grounded in the general sequence of the progression of skills and appropriate developmental concepts. Secondary school principals expressed this need less precisely, describing their activity as "judging the whole." Principals at both levels articulated a philosophy of evaluation grounded in one or more fundamental principles to which they **individually** were committed as educators. But, it was abundantly clear that whatever was identified by principals as unique about their school, that uniqueness was within the

prescribed reality of what was acceptable to state and district guidelines. Thus, while acknowledging differences among schools, the principals used the full range of qualitative measures summarized in Table 1 to make further judgments within the range of already established goals and objectives. In other words, there was no indication of critical judgments of curricular goals outside the prescribed boundaries.

Interactions between Decision-Making and Negotiations

While the public equates evaluation with control, principals readily admit that they do not and cannot control all school variables. Therefore, they substitute social processes for control (Bogotch, 1990). Principals "encourage and coax." "If you stop an activity, you are liable to stop all activities," said one of the elementary principals. These principals perceived that the public feels that the end of the evaluation process should be either to get rid of bad teachers or to affirm that good teaching is occurring. Time and again, principals voiced their opposition to this definition of evaluation, replacing it with growth statements filled with improvement-orient concepts and contextual explanations.

Nonetheless, there is a gap between what is and what ought to be, and between what ought to be and what the public believes should occur. Ideally, principals noted, "we work with teachers as a process to provide opportunities for professional growth. We do that to a degree. I wish we did more." Added another, "What we do is to help teachers get better. This is not viewed

by the public as part of evaluation."

Principals stressed the actions taken to foster professional growth. They "researched" problems, identified resources to help teachers, "found practical examples where it was working," linked teachers together in teacher to teacher models, and documented problems. At times, they admitted to "overlooking short-term problems." They are willing to go along with teachers who show potential or who are having personal problems. They also give time for teachers to develop and they show patience. Principals talked about how they give teachers "free" visits before formal classroom observations and how they balance positive and negative comments. There were numerous statements attesting to the practice that principals deliberately withhold negative comments from teachers.

Although the terms fairness and honesty were mentioned as important in decision-making, the principals were aware of the interrelationship between themselves and staff. "I tell people whom I am closest to that I will be harder on them. That's not fair, but it's how I counterbalance...." "We all have biases. Hopefully, I'm capable and open enough to be fair." "You may make allowances in the short-term; but over the long haul, you have a responsibility [to the children and to the school system]...you might as well keep my salary if I do not perform my duties." Nevertheless, "fairness may not always be perceived by others."

"I will accept the criticism that we all try to help

individuals longer than we should." "The dynamic in business is that if you do not do the job, you get fired. We do not do this in education. We must share that responsibility." "In 11 years, I have recommended only one person for termination." But, each of the principals mentioned that they have counselled numbers of teachers into other careers. "Teachers know when they are not successful."

True or not, one of the most prevalent personnel strategies within schools is the transferring of teachers within the school system. Ironically, when one principal mentioned that he, too, engaged in this practice, a colleague on the panel said, "Yes, I know. I got one of them." One of the reasons why this unsatisfactory procedure endures is that terminating a teacher is difficult. Principals expressed here and elsewhere (Bogotch, 1990) the feeling of being put on trial at a termination hearing. Even with complete documentation and the full due process protection of the teacher's constitutional rights, they felt as if it was "their word against his." Furthermore, those responsible for making the ultimate decision, have not always followed through. After awhile, "you get 'gun shy.'" "At these times, it is unfortunate that there are lovely, warm people who love children, but cannot teach." We see that principals take their time and seem reluctant to take final action. Once they do make their decision, however, it is up to educators at other levels and outside the building to go beyond the warm, human qualities of the person and reach a decision about teaching."

"Everyday, we are constantly evaluating what is going on, how can I do this better. When I say, 'I don't like thus and such,' we go into an evaluative mode. Not a day goes by that I don't say this." While the principals agree that their leadership impacts on what happens within the schools, they believe that it is "terribly unfair to say that we [i.e., schools] are responsible for the ills of society. There are so many variables in education; it's really difficult...we [i.e., schools] are doing a damn fine job."

Discussion

The data from these principals provide a basis for analyzing the processes and indicators used in their daily practice. Principals recognize that evaluation while essentially subjective involves trying to understand on a daily basis the total school context. Although there are gaps in terms of the logic and coherence among the processes and indicators, by Dewey's standards, we found little evidence of final judgments and opinion-centered evaluation. As to impressionistic judgments, however, it is clear that principals make judgments and decisions based on extremely little information. It is not just that the concepts are difficult to articulate and intangible; it is still an unexamined empirical question as to the amount of analysis and synthesis that goes into the making of evaluative judgments.

Much of the data presented here falls outside of the normative evaluation models such that practitioner behaviors are open to criticism as not being objective, systematic, goal-

oriented, rational, or negotiated. The behavioral data based on perceptions described by these practitioners is different from the evaluation specialists' literature. At the same time, however, their behaviors fall short of Dewey's critical judgment framework in that the judgments are not always based on strength; they do not fully support that which is good nor make what works more secure nor readily eliminate what is detrimental. While critical of inappropriate measures, goals, and decisions, the principals in this study did not express strategies for challenging the public's non-growth, non-educational realities. At times, they even voiced their support of measures stated to be irrelevant to learning, describing the time and effort devoted to improving the standardized test scores. This time and effort, of course, must be taken from the other stated goals of diversity, critical thinking, self-esteem, social activities, and sports.

What emerges is a contradictory reality. On the one hand, principals state that they do not believe that twice a year formal evaluations tell them very much about a teacher's ability. As one principal stated, "you can't measure good teaching with a formal instrument." Another said, it has "no impact whatsoever." Yet, it was clear that if a teacher could not **perform** for evaluators twice a year in these packaged segments, then "we have a problem." Such statements tend to widen the gap between the evaluation specialist who designs the rules and behaviors and the principal who overlays his or her own reality on top of the process. As a result, Dewey's criterion of making what is good,

more secure, more coherent is undermined, however unintentionally. Both through words and actions, principals contribute to these opposite effects.

Another way of looking at these contradictory realities is that the evaluation specialists establish a research-based, objective and comprehensive model of competent teaching. The observation procedures are spelled out as clearly as possible in order to achieve "objectivity." Yet, all principals view classrooms within a larger, continuous context, taking into account the history of the teacher and the interactions among students and teachers. A vivid example of this contradiction was expressed by one principal in this study. The day before a formal observation, a teacher hit a child with a ruler. The principal spoke with the teacher immediately afterwards. During the formal observation, the principal noted that this same teacher was carrying a ruler in her hand and that students in the front rows were flinching as the teacher punctuated the lesson. Based on prior knowledge, the principal gave the teacher a low evaluation. On the other hand, two "outside" observers recording only the observable teaching behaviors listed on the formal instrument, gave the teacher positive evaluations. When all three scores were averaged together, the teacher was rated as adequate, thus, negating the principal's assessment. Of course, the principal has other avenues beside the formal teacher evaluation process to pursue such matters, but the point is that the total picture cannot be attained by "objective" and

quantified measures.

There is another aspect of formal teacher evaluation instruments which also tends to widen the gap between the evaluation specialist and practitioner. Principals are not likely to give equal weight to each item on the checklist. In fact, all of the principals in this study and previously studied by Bogotch (1990) responded that they inform teachers of their priorities, whether it be higher order learning skills, writing, classroom management, group work, global education, etc. The message sent to teachers during formal evaluations is that they are to perform differently from their normal way of teaching and are permitted to deviate from the school's instructional priorities. Again, such messages make people feel less secure in judging and deciding the worth of what they regularly do.

The data also raises serious doubts about principals' roles **vis a vis** curriculum. On the one hand, the issue is discussed in terms of technical expertise; but, on a deeper, more significant level, the question is whether principals have the knowledge-base to critically examine what is being taught in terms of the overall goals for the school. All of the principals avoided direct responses to curricular issues. They discussed school-wide goals, instructional leadership, but not the appropriateness of the goals, measures, or fit between what is being taught and the goals. All this falls, they say, under the rubric of teacher trust. But they know that teachers are following state and district guidelines and that without school restructuring

support, teachers are not in a better position than principals to effect school-wide or system-wide changes in curriculum. It ultimately comes down to a matter of leadership. The question is what role will school principals play in future curricular reform debates. At the school level, a laissez-faire attitude towards the current subject matter would be as disturbing to Dewey and Whitehead today as it was when they were writing.

The issue of principal-teacher trust is also unclear. In a previous study on discretionary behaviors, principals referred to certain school activities as occurring "naturally" (Bogotch, 1990) Yet, when teachers were subsequently asked whether these activities actually occurred, they often "didn't know." The point is that principals choose which areas to activity manage and which to ignore. Labelling the latter as trust is at best misleading, at worst, irresponsible and untrue. The logic of confidence may be something other than a professional model. As we continue to learn about the logic of experience, however, we should be able to give words to what constitutes a professional educational model.

Moreover, negotiated evaluation models explicitly put responsible parties on an equal footing; it is hard to interpret the level of equality from the data here. One of the more insightful statements equated evaluation with communication. "We are constantly communicating; maybe that's what evaluation is - communicating with people who are responsible - communication between the responsible parties involved to make things better.

It is a part of our inherent humanness." But in almost the very next sentence that same principal states that "perhaps feedback is discounted; that's why change doesn't occur." If the evaluation processes were truly negotiable, then feedback would not be discounted.

None of the principals in this study offered specific recommendations for policymakers or the public. What they said was that there were public misconceptions and inordinate and unfair expectations about their work as it relates to evaluation. They felt that the 'state of the art' in evaluation is insufficient in terms of the complexity of educational issues. That is, neither standardized achievement tests nor formal instruments and procedures capture the totality of the experiences needed to judge students, teachers, or curricula. Quantified achievement test scores and checklist teaching behaviors offer the unregenerative public the opportunity to declare results and make final judgements. We recognize how far practitioners need to go to attain Dewey's criteria of critical judgments; nevertheless, on balance, data from these practitioners are much closer to Dewey's framework than to the normative models of the evaluation specialists.

Recommendations

That evaluation now occurs haphazardly and informally does not mean that it cannot be conducted more systematically and improved. School professionals need a logic of experience. Towards that objective, we feel that increased professional

scrutiny and public testing would improve what principals do in evaluating programs and personnel. Systematically developing reflective practice processes which can be explained and verified is essential in winning support from both the public and policymakers.

One of the first steps in developing reflective knowledge is to increase collegial communications and interactions among principals, teachers, and evaluation specialists. Although the vocabulary for this new action evaluation science is underdeveloped, the on-going practices and value-orientations shared by these principals are, for the most part, verifiable. The first step is for practitioners and evaluation specialists need to work towards the collaborative development of linguistic conceptions of evaluation as it is practiced.

The long list of qualitative evaluation indicators in Table 1 captures the essence of what school principals care about most. The indicators are not behavioral checklists, but rather measures for judgments and decisions. The current "state of the art" indicators raise for us at least three concerns: (1) that qualitative measures are unsystematic, (2) that judgments are based on limited information, and (3) that evaluation is too complex to be left to any one person's perspective. Systematic analysis is important in order to advance knowledge beyond the surface level of descriptions. The qualitative indicators reported here go beyond externally directed checklists in that they integrate behaviors with judgments. Thus, the validation

task involves systematically examining behaviors, judgments, and their interactions in order to solve the mystery of discretionary evaluations.

The second concern is that principals appear to be making evaluative judgments based on very limited information. Standing outside a classroom, having a short conversation with a student or parent, glancing at a list of teacher referrals, etc. is not sufficient to meet the kind of discriminating and unifying analysis and synthesis of critical judgments. Some principals are better than others at using qualitative measures for school improvement (Bogotch, 1990), but since it is left to each principal individually to learn these skills, the process does not instill public confidence nor meet accountability demands.

The third concern is that the complexity of educational realities and negotiated evaluation is actually beyond the limits of any one individual alone. Evaluation is critical for school improvement and merits input from as many participants as is feasible. In non-routine matters, collective judgment is superior to individual judgment, not in terms of "objectivity," efficiency, or final judgements, but rather, in terms of total school effectiveness. However educational systems evolve, they must always allow for individual differences which are often beyond the experiences and judgments of any one person.

Another important recommendation which emerges from this study is that principals need central office support. The principals in this sample readily admitted that they often invest

too much time on certain individuals and programs. Obviously, the time expended varies with every situation, but extending time-lines beyond reasonable limits is less likely to occur if the larger system supported solidly documented school practitioner decisions. Too often, actions by boards and the central office constrain principals' decision-making arbitrarily or capriciously. Terminations are a good example. When established criteria are followed in good faith, it is essential that principals be supported. Failure to do so causes principals to feel as if they are on trial, resulting in their becoming "gun shy," not only in matters dealing with personnel, but in other areas of evaluation as well.

Principal decision-making, however, is not a straightforward issue. A number of Principal Assessment models include decisiveness as one school principal competency (e.g., FCEM; NASSP). This sends a clear, but misleading message to principals. That is, the quality of the decision is less important than the decision itself. Theoretically and practically there is little support for this competency. According to Drucker (1966), effective managers make very few decisions, choosing instead to focus on important ones. In fact, there is really no decision until something is put into practice. In terms of Dewey's framework and the statements made by the principals here, patience, purpose, and growth, rather than decisiveness are essential. The multiple variables to be considered by educators are in direct conflict with "the pressure

to demonstrate immediate utility in school administration and instruction..." (Dewey, 1929, p.23).

It is likely that those principals who view themselves as professional are more able to instill professionalism in others. There is evidence from school effectiveness literature that curricular-oriented principals care about the quality of academic programs and transmit professionalism throughout the school. The problem is that the research literature does not investigate professional practice or decipher the "patterned relationships between actor" (Bacharach & Lawler, 1982, p. 7). Thus, the findings of qualitative measures and discretionary behaviors as reported here are not reflected or translated into policies, public confidence and trust, nor do they add to the body of knowledge which informs professional practice. Said another way, private, inconsistent, and idiosyncratic uses of discretionary methods will not bring credibility to the profession. This is one of the major challenges facing our schools. The public and policymakers' preoccupation with standardized measures is contrary to quality evaluation procedures and learning theories. Practitioners will need the help of educational specialists in developing discretionary criteria and procedures of evaluation.

References

Bacharach, S., & Lawler, E. (1982). Power and politics in organizations. San Francisco: Jossey-Bass.

Bailyn, B. (1960). Education in the forming of American society. New York: W.W. Norton & Co.

Berke, I. & Law, A. (1981). Prospects for educational evaluation. In Cirincione-Coles (Ed.). The future of education: Policy issues and challenges. Beverly Hills: Sage Publications.

Bogotch, I. (1990). A model of school managerial control: The systematic analysis of processes, behaviors, and indicators. Unpublished dissertation Florida International University

Cronbach, L. (1963). Course improvement through evaluation. Teachers College Record, 64(8), 672-686

Dewey, J. (1939/1970). Philosophy, education, and reflective thinking. In T. Buford (Ed.). Essays on other minds. Urbana: University of Illinois Press. Previously published in J. Ratner (Ed). Intelligence and the Modern World: John Dewey's philosophy. New York: The Modern Library.

Dewey, J. (1929). The sources of a science of education. New York: Liveright, p. 23.

Drucker, P. (1966). The effective executive. New York: Harper & Row.

Elliot, R. (1972) Reexamining the notion of the universality of educational measurement. Paper presented at the Annual Meeting of the National Council for the Social Studies. Boston.

Guba, E. & Lincoln, Y. (1989). Fourth generation evaluation. Newbury Park: Sage Publications, Inc.

McDermott, R. & Tylbor, H. (1987). On the necessity of collusion in conversation. In L. Kader (Ed.). Power through discourse. Norwood, NJ:Ablex Publishing Corp. 153-170.

Mintzberg, H. (1971). Managerial work: Analysis from observation. Management Science, 18(2) 97-100.

Mintzberg, H., Raisinghansi, & Theoret. (1976). The structure of 'unstructured' decision processes. Administrative Science Quarterly, 21 246-275.

Popham, J. (1988). Educational evaluation. Engelwood Cliffs: Prentice Hall.

Ryle, G. (1969). The concept of mind. New York: Barnes & Noble, Inc.

Scriven, M. (1967). Methodology of evaluation. In R. Stake (Ed.). Perspectives in curricular evaluation. American Educational Research Association Monograph Series on Evaluation, No. 1. Chicago: Rand McNally.

Simon, H. (1946). The proverbs of administration. Public Administration Review, 6, 53-67.

Stake, R. (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540.

The Joint Committee on Standards for Educational Evaluation (1988). The Personnel Evaluation Standards. Newbury Park: Sage Publications.

The Joint Committee on Standards for Educational Evaluation (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill Book Company.

Vickers, G. (1967). Towards a sociology of management. New York: Basic Books.

Whitehead, A. (1951). The aims of education. New York: Mentor Books.

Williams, P. (1991). Quality control of curriculum and instruction in the elementary school. Unpublished dissertation, Florida International University